Perturbation theory for the tensor rank decomposition

Nick Vannieuwenhoven

FWO / KU Leuven

January 18, 2016

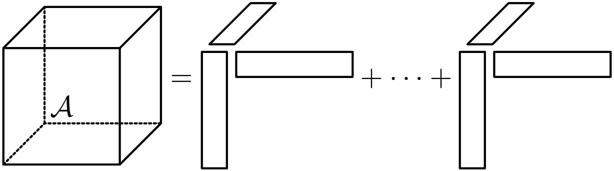
Overview

Overview

Tensor rank decomposition

Hitchcock (1927) introduced the tensor rank decomposition:[1]

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$$



The rank of a tensor is the minimum number of rank-1 tensors of which it is a linear combination.

[1] Candecomp, Parafac, Canonical polyiadic, or CP decomposition.

Identifiability

A rank-1 tensor is uniquely determined up to scaling:

$$\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} = (\alpha \mathbf{a}) \otimes (\beta \mathbf{b}) \otimes (\alpha^{-1}\beta^{-1}\mathbf{c}).$$

Kruskal (1977) proved that the rank-1 terms appearing in

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d$$

are uniquely determined if r is small and $d \geq 3$.

Generic identifiability

It is expected[2] by [BCO13, COV14] that a random complex tensor rank decomposition

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d,$$

of strictly subgeneric rank, i.e.,

$$r \leq \left\lceil \frac{n_1 n_2 \cdots n_d}{n_1 + \cdots + n_d - d + 1} \right\rceil - 1,$$

is identifiable with probability 1, provided that it is not one of the exceptional cases where $(n_1, n_2, \ldots, n_d)$ is

- (n_1, n_2), or
- $(4, 4, 3)$, $(4, 4, 4)$, $(6, 6, 3)$, $(n, n, 2, 2)$, $(2, 2, 2, 2, 2)$, or
- $n_1 > \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1)$ (unbalanced).

---

[2] [COV14] proved the conjecture when $n_1 n_2 \cdots n_d \leq 17500$.

Generic symmetric identifiability

It was proved by [COV16] that a random complex symmetric tensor rank decomposition

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i \otimes \mathbf{a}_i \otimes \cdots \otimes \mathbf{a}_i,$$

of strictly subgeneric rank, i.e.,

$$r \leq \left\lceil \frac{\binom{n-1+d}{d}}{n} \right\rceil - 1,$$

is identifiable with probability 1, provided that the space is not one of the following exceptional cases:

- symmetric matrices, i.e., $S^2\mathbb{C}^n$,
- $S^3\mathbb{C}^6$,
- $S^4\mathbb{C}^4$, or
- $S^6\mathbb{C}^3$.

Overview

Perturbations and conditioning

In applications it is uncommon to work with the “true” tensor $\mathcal{P}$. Usually we only have some $\widehat{\mathcal{P}}$ for which

$$\|\mathcal{P} - \widehat{\mathcal{P}}\| = \epsilon$$

with ϵ small.

The error can originate from many sources:

- measurement errors,
- model errors, and
- accumulation of round-off errors.

Perturbations and conditioning

A true decomposition

$$\mathcal{P} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d$$

is nice, but I only know $\widehat{\mathcal{P}}$. I can compute an approximation

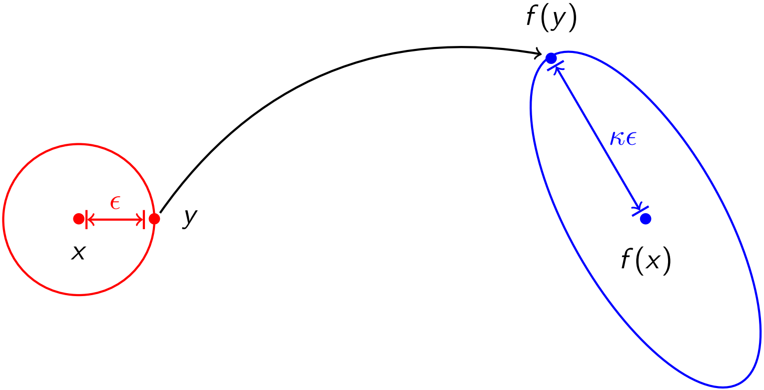
$$\widehat{\mathcal{P}} \approx \sum_{i=1}^{r} \widehat{\mathbf{a}}_i^1 \otimes \widehat{\mathbf{a}}_2^d \otimes \cdots \otimes \widehat{\mathbf{a}}_i^d \approx \mathcal{P}$$

but what does it tell me about $\mathcal{P}$?

- Is $\mathcal{P}$'s decomposition unique?
- Are the terms in $\widehat{\mathcal{P}}$'s decomposition related to those of $\mathcal{P}$?
- Can I find an upper bound on this difference?

Condition number

The condition number gives the first-order answer to the question: How sensitive is f to perturbations to the input?



Condition number

Definition

The *relative condition number* of a function $f : X \to Y$ at $x \in X$ is

$$\kappa = \lim_{\epsilon \to 0} \max_{\|\Delta x\|_\beta = \epsilon} \frac{\|f(x) - f(x + \Delta x)\|_\alpha / \|f(x)\|_\alpha}{\|\Delta x\|_\beta / \|x\|_\beta},$$

for some norms $\| \cdot \|_\alpha$ and $\| \cdot \|_\beta$.

Condition number

Definition

The *relative condition number* of a function $f : X \to Y$ at $x \in X$ is

$$\kappa = \lim_{\epsilon \to 0} \max_{\|\Delta x\|_\beta = \epsilon} \frac{\|f(x) - f(x + \Delta x)\|_\alpha / \|f(x)\|_\alpha}{\|\Delta x\|_\beta / \|x\|_\beta},$$

for some norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$.

The condition number is the property of a *problem*, not of a particular algorithm solving the problem.

Overview

An experiment in Tensorlab

Let $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5\end{bmatrix}$, $B = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 & \mathbf{b}_5\end{bmatrix}$, and

$$C_\epsilon = \begin{bmatrix}\mathbf{c} + \epsilon\mathbf{c}_1 & \mathbf{c} + \epsilon\mathbf{c}_2 & \mathbf{c} + \epsilon\mathbf{c}_3 & \mathbf{c} + \epsilon\mathbf{c}_3 & \mathbf{c} + \epsilon\mathbf{c}_5\end{bmatrix}.$$

be 5×5 matrices with $\mathbf{a}_i$, $\mathbf{b}_i$, and $\mathbf{c}_i$ random vectors.

Consider a sequence of tensors

$$\mathcal{T}_\epsilon = \sum_{i=1}^{5} \mathbf{a}_i \otimes \mathbf{b}_i \otimes (\mathbf{c} + \epsilon\mathbf{c}_i) \quad \xrightarrow{\epsilon\to 0} \quad \sum_{i=1}^{5} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}.$$

Then,

- $\mathcal{T}_\epsilon$ is 5-identifiable if $\epsilon \neq 0$, while
- $\mathcal{T}_0$ has ∞-many decompositions.

An experiment in Tensorlab

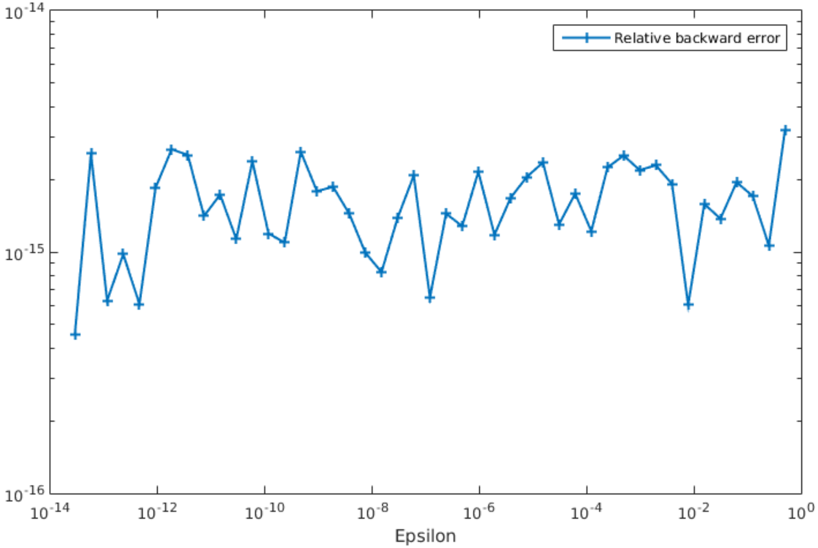
Let's compute the unique decomposition of $\mathcal{T}_\epsilon$ in Tensorlab using an algebraic algorithm:

```
% Generate tensor.
T_eps = cpdgen({A,B,C_eps});

% Compute decomposition.
[U, out] = cpd_gevd(T_eps,5);

% Measure relative backward error.
relBackErr = frob(T_eps - cpdgen(U)) / frob(T_eps);
```

An experiment in Tensorlab



An experiment in Tensorlab

A small backward error is but the beginning in a data-analysis application. We need a small forward error!

Let $\widehat{A}_\epsilon$, $\widehat{B}_\epsilon$, $\widehat{C}_\epsilon$ be the computed factor matrices of $\mathcal{T}_\epsilon$. Then, we really seek small errors

$$\frac{\|A - \widehat{A}_\epsilon P\|}{\|A\|}, \quad \frac{\|B - \widehat{B}_\epsilon P\|}{\|B\|}, \quad \text{and} \quad \frac{\|C_\epsilon - \widehat{C}_\epsilon P\|}{\|C\|},$$

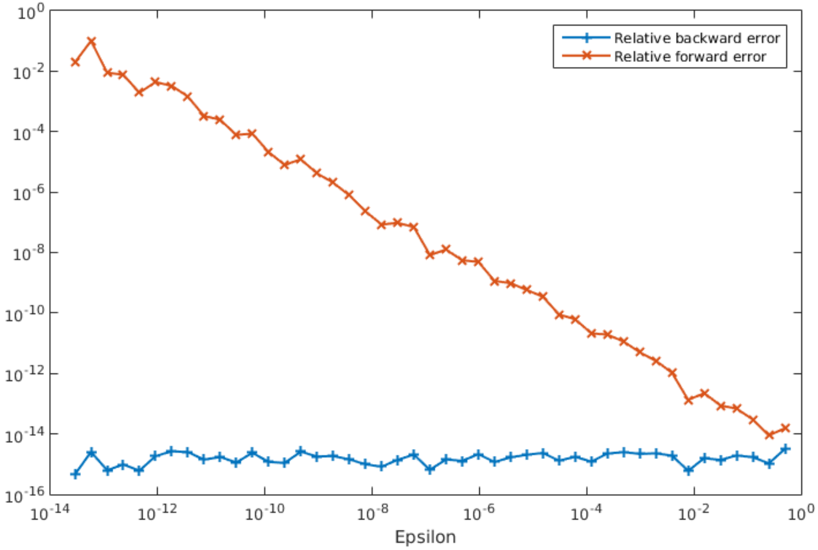
for some permutation matrix P.

An experiment in Tensorlab

In our toy problem, we can compute the forward error using:

```
% [U, out] = cpd_gevd(T_eps,5);
U_true = {A,B,C_eps};
errs = cpderr(U_true,U);
```

An experiment in Tensorlab



An experiment in Tensorlab

Except for toy problems, we don't know the true decomposition!
So how do we assess the forward error??

An experiment in Tensorlab

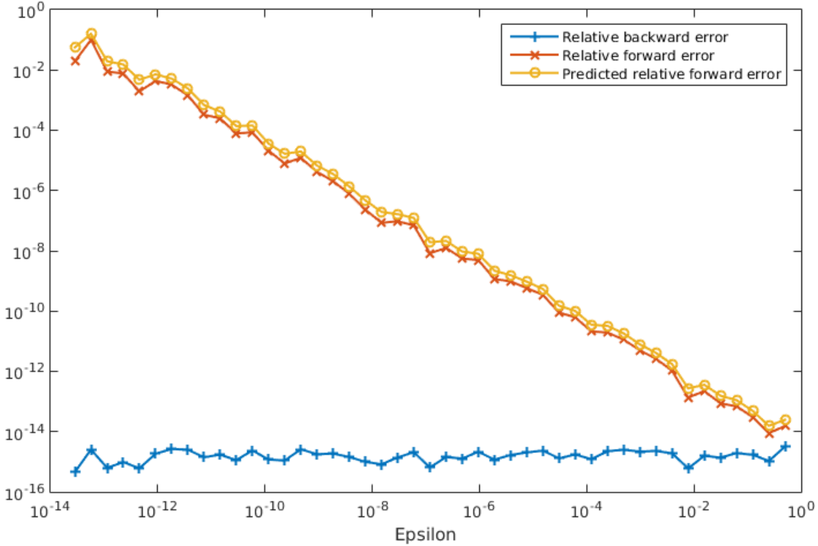
Except for toy problems, we don't know the true decomposition!
So how do we assess the forward error??

The condition number

By the definition of the condition number, we have:

$$\text{(relative) forward error} \approx \text{(relative) backward error} \times \text{(relative) condition number.}$$

An experiment in Tensorlab



Overview

Terracini's matrix

For every

$$p_i = \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathbb{C}^{n_1 n_2 \cdots n_d},$$

we define the matrix

$$T_i = \begin{bmatrix} I_{n_1} \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d & \cdots & \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^{d-1} \otimes I_{n_d} \end{bmatrix}.$$

Then, Terracini's matrix is given by

$$T_{p_1,\ldots,p_r} = \begin{bmatrix} T_1 & T_2 & \cdots & T_r \end{bmatrix}.$$

Terracini's matrix is singular

Working in affine space,

Terracini's matrix is never of maximal rank $r(n_1 + n_2 + \cdots + n_d)$.

Let us write the singular values of $T_{p_1,\ldots,p_r}$ as

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0 \geq 0 \geq \cdots \geq 0$$

with $N = r(n_1 + \cdots + n_d - d + 1)$. Then, I claim that

$$\sigma_N^{-1}$$

may be interpreted as an absolute condition number.

A condition number

Theorem (V, 2016)

Let $T_{p_1,\ldots,p_r}$ be Terracini's matrix associated with the rank-1 tensors $p_i \in \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$. Let $N = r(n_1 + \cdots + n_d - d + 1)$. Then,

$$\kappa_A = \sigma_N^{-1}$$

is an absolute condition number of the rank decomposition problem at $\mathcal{A} = p_1 + p_2 + \cdots + p_r$.

Overview

A well-conditioned example

Let's take the "Amino acids" data set from `http://www.models.life.ku.dk/nwaydata`, which is "a nice set for fitting PARAFAC."

> *"The data set consists of five simple laboratory-made samples. Each sample contains different amounts of tyrosine, tryptophan and phenylalanine dissolved in phosphate buffered water. The samples were measured by fluorescence (excitation 250-300 nm, emission 250-450 nm, 1 nm intervals) on a spectrofluorometer. The array to be decomposed is hence* $5 \times 51 \times 201$. *Ideally, these data should be describable with three PARAFAC components."*

A well-conditioned example

Armed with the condition number, we can quantify precisely how "nice" this tensor is.

```
>> load 'real_data/amino.mat'; T = X;

% Compute approximate decomposition.
>> U = cpd(T,3);

% Compute condition number.
>> condNumber = cpdcondest(U,'relative')
ans = 3.1159e+00
```

The condition number is $\mathcal{O}(1)$, so the backward error is hardly amplified!

A well-conditioned example

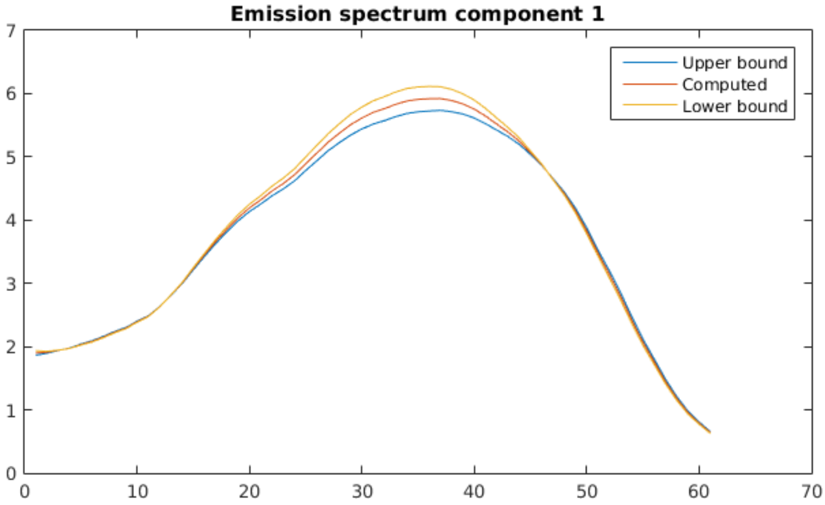
```
% Compute backward error.
>> relBErr = frob(T - cpdgen(U))/frob(T);
ans = 2.5049e-02

% Estimate forward error.
>> relFErr = relBErr * condNumber
ans = 7.8049e-02
```

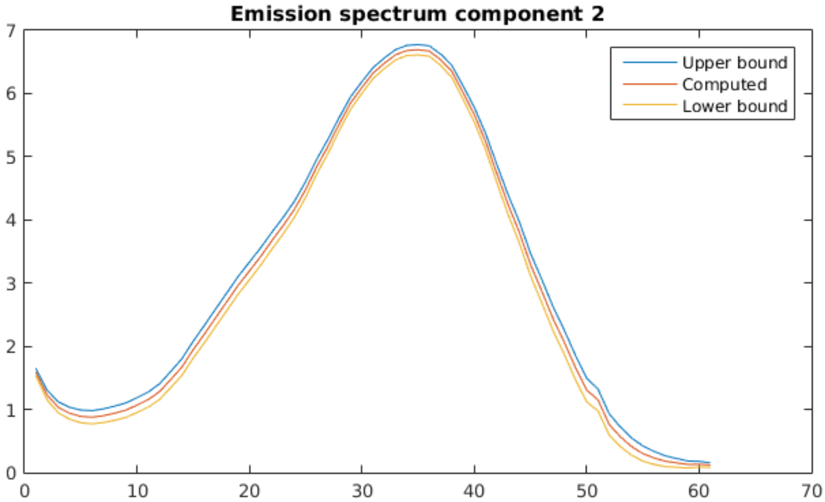
If you can bound the measurement errors by `relMErr`, then one can even obtain the forward error w.r.t. the true decomposition:

```
% Estimate forward error w.r.t. true decomposition.
>> relFErr = (relBErr + relMErr) * condNumber
```

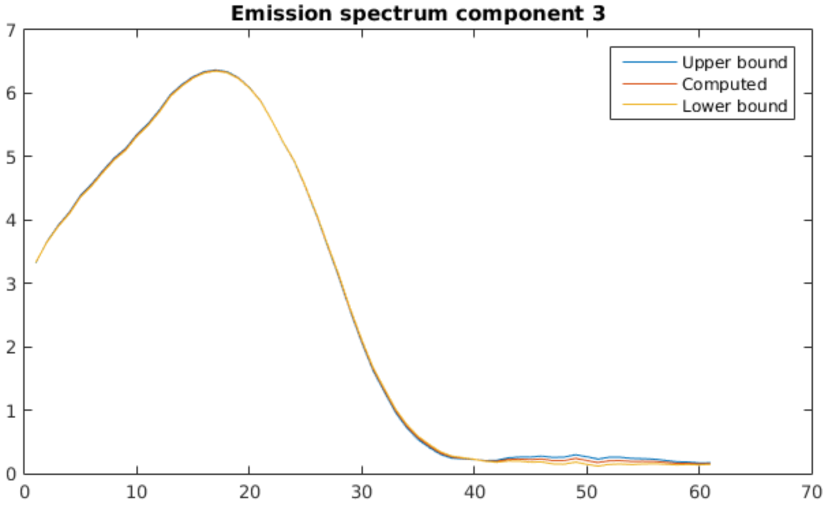
A well-conditioned example



A well-conditioned example



A well-conditioned example



Ill-posedness and ill-conditioning

The classic example from [dSL08] is the rank-3 tensor

$$\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u},$$

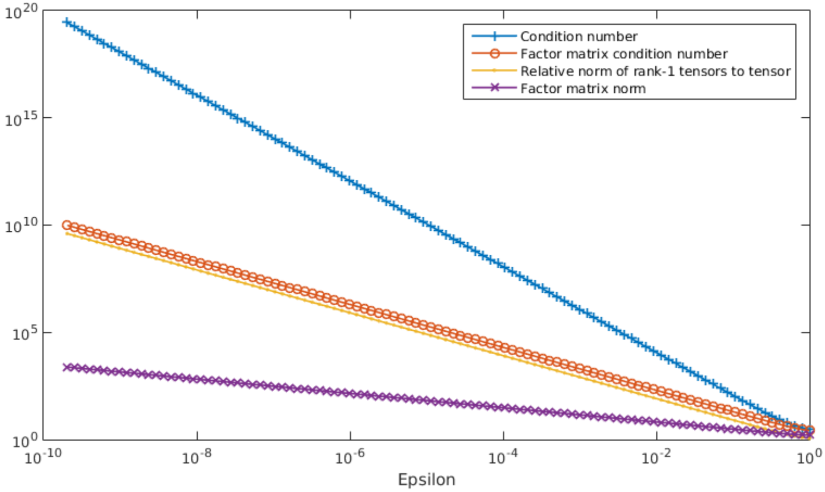
which is a limit of identifiable rank-2 tensors:

$$\lim_{\epsilon \to 0} \left(\frac{1}{\epsilon} \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u} - \frac{1}{\epsilon} (\mathbf{u} + \epsilon \mathbf{v}) \otimes (\mathbf{u} + \epsilon \mathbf{v}) \otimes (\mathbf{u} + \epsilon \mathbf{v}) \right).$$

[V16] proves that "a sequence of diverging components" entails that the relative condition number $\kappa \to \infty$.

Formally: As you move towards an open part of the boundary of the r-secant variety of a Segre variety, the relative condition number becomes unbounded.

Ill-posedness and ill-conditioning



Overview

Conclusions

Take-away messages:

- Tensors are conjectured to be identifiable [COV14].
- Symmetric tensors were proved to be identifiable [COV16].
- Forward errors matter.
- The condition number multiplied with the backward error bounds the forward error "to first order."
- The condition number of a decomposition can be computed practically.

Thank you for your attention!

Further reading

Main references:

- Vannieuwenhoven, *A condition number for the tensor rank decomposition*, 2016. (In preparation)
- Chiantini, Ottaviani, and V, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIAM J. Matrix Anal. Appl., 2014.

References

General introduction

- [H1927] Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 1927.
- [K1977] Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Lin. Alg. Appl., 1977.

Conditioning

- [BC13] Bürgisser and Cucker, *Condition: The Geometry of Numerical Algorithms*, Springer, 2013.
- [dSL08] de Silva and Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 2008.
- [V16] Vannieuwenhoven, *A condition number for the tensor rank decomposition*. (In preparation)

References

Generic identifiability

- [BCO13] Bocci, Chiantini, and Ottaviani, *Refined methods for the identifiability of tensors*, Ann. Mat. Pura Appl., 2013.
- [CO12] Chiantini and Ottaviani, *On generic identifiability of 3-tensors of small rank*, SIAM J. Matrix Anal. Appl., 2013.
- [COV14] Chiantini, Ottaviani, and Vannieuwenhoven, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIAM J. Matrix Anal. Appl., 2014.
- [COV16] Chiantini, Ottaviani, and Vannieuwenhoven, *On generic identifiability of symmetric tensors of subgeneric rank*, Trans. Amer. Math. Soc., 2016. (Accepted)
- [DdL15] Domanov and De Lathauwer, *Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL*, SIAM J. Matrix Anal. Appl., 2015.
- [HOOS15] Hauenstein, Oeding, Ottaviani, and Sommese, *Homotopy techniques for tensor decomposition and perfect identifiability*, arXiv, 2015.